Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 6803–6808, June 1997
Evolution

Ligand binding was acquired during evolution of nuclear receptors

HECTOR ESCRIVA*, RACHID SAFI†, CATHERINE HÄNNI†, MARIE-CLAIRE LANGLOIS†, PIERRE SAUMITOU-LAPRADE‡, DOMINIQUE STEHELIN†, ANDRÉ CAPRON*, RAYMOND PIERCE*, AND VINCENT LAUDET†§

†Centre National de la Recherche Scientifique Unite Mixte de Recherche 319, Mécanismes du Développement et de la Cancérisation, and *Institut National de la Santé et de la Recherche Médicale U167 Relations Hôte-Parasite et Stratégies Vaccinales, Institut Pasteur de Lille, 1 rue Calmette, 59019, Lille Cedex, France; and ‡Centre National de la Recherche Scientifique Unite de Recherche Associée 1185 Génétique et Evolution des Populations Végétales, Université de Lille 1, Bâtiment SN2, 59655, Villeneuve d'Ascq, France

Communicated by Roy J. Britten, California Institute of Technology, Corona Del Mar, CA, April 25, 1997 (received for review December 5, 1996)

ABSTRACT **The nuclear receptor (NR) superfamily comprises, in addition to ligand-activated transcription factors, members for which no ligand has been identified to date. We demonstrate that orphan receptors are randomly distributed in the evolutionary tree and that there is no relationship between the position of a given liganded receptor in the tree and the chemical nature of its ligand. NRs are specific to metazoans, as revealed by a screen of NR-related sequences in early- and non-metazoan organisms. The analysis of the NR gene duplication pattern during the evolution of metazoans shows that the present NR diversity arose from two waves of gene duplications. Strikingly, our results suggest that the ancestral NR was an orphan receptor that acquired ligand-binding ability during subsequent evolution.**

The nuclear receptor (NR) superfamily contains ligand-activated transcription factors that exert widely different biological responses by regulating target gene expression positively and/or negatively (1, 2). Apart from receptors that bind steroid hormones, retinoic acid, or thyroid hormones, this superfamily contains so-called orphan receptors for which no ligand is known. Recently, ligands have been proposed for some orphan receptors such as PPAR, LXR, or FXR. Interestingly most of these newly discovered ligands act in an intracrine fashion. However, in most cases, and despite intensive efforts, ligands have not been found for most orphan receptors (e.g., COUP-TF, HNF4, SF1, etc.), the biological role of which remains enigmatic (1, 2).

The modular organization of NRs, the various degrees of conservation between their respective domains and the importance of NRs for many physiological processes in both arthropods and vertebrates, have led several authors to study these molecules from an evolutionary point of view (3–5). We showed previously that the NR superfamily is ancient in origin and was well diversified before the arthropod/vertebrate split, because most of the subfamilies of NRs that we defined, and most of the groups of receptors, were already present before this divergence (3). However, the precise origin of the family was still unclear. To elucidate this crucial point, the origin and diversification of NRs was studied by amplifying fragments of NR genes in various organisms that are located at critical positions in the evolutionary tree. Our analysis demonstrates that NRs appear specific to metazoans and that two waves of gene duplications led to their present diversity. Furthermore, our results suggest that orphan receptors are ancient and that liganded receptors independently acquired the ability to bind a ligand later in evolution.

Table 1. Authenticity of the PCR products

Clone	%AT	Southern blot	Genomic/cDNA
Hydra COUP	+		C
Schistosoma TR4	+	+	G
Schistosoma RXR	+	+	GC
Schistosoma FTZ-F1	+	+	C
Schistosoma COUP	+	+	
Schistosoma COUPII	+	+	
Phallusia FTZ-F1			G
Branchiostoma COUP			C
Branchiostoma TR2			C
Danio REVERB			C
Danio COUP			C
Danio PPAR			C
Danio RAR			C
Danio NGFIB			C

%A/T +: Sequence of the PCR fragment is rich in bases A and T, as is the genome of the corresponding organism. Southern blot +: Positive signal with the corresponding probe under high stringency conditions. G: Isolation of a genomic clone. C: Isolation of cDNA clones.

MATERIALS AND METHODS

DNA Extraction. DNA was extracted according to standard procedures (6) from zebrafish (*Danio rerio*), amphioxus (*Branchiostoma lanceolatum*), acorn worm (*Balanoglossus* sp.), sea urchin (*Sphaerechinus granularis*), tunicate (*Phallusia mamillata*), mollusk (*Biomphalaria glabrata*), annelid (*Perinereis cultrifera*), brine shrimp (*Artemia salina*), flatworm (*Schistosoma mansoni*), cnidarians (*Hydra vulgaris* and *Anemonia sulcata*) fresh tissues; and from shark (*Scyliorhinus canicula*), lamprey (*Petromyzon marinus*), hagfish (*Myxine glutinosa*) frozen tissues. Sponge (*Pseudocorticium jarre*) DNA was a gift of R. Garrone (Lyon, France). Non-metazoan samples were from plants (*Beta vulgaris*, *Thymus vulgaris*, *Daucus carota*, and *Sparganium erectum*), a fungus (*Glomus intraradices*), a red alga (*Gracilia verrucosa*), and a protist (*Trypanosoma cruzi*).

PCR Analysis. Three or four primers were designed for each studied group of receptors (TR, RAR, PPAR, REV-ERB, ROR, EcR, LXR, NGFIB, ER, steroid receptors, FTZ-F1, RXR, COUP-TF) and all possible combinations of primers were used in semi-nested PCR amplifications. Primer sequences are available from V.L. upon request. PCR experiments were done according to a modified "touch-down" method (7) using the *Taq* polymerase and buffer from Eurogentec (Brussels). DNA extraction, PCR amplification, and cloning/sequencing were performed under strict conditions to avoid contamination. All the PCR experiments were carried out in the laboratory of P.S.-L., where no animal DNA is

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: NR, nuclear receptor.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. U93406–U93489).

§To whom reprint requests should be addressed. e-mail: vlaudet@infobiogen.fr.

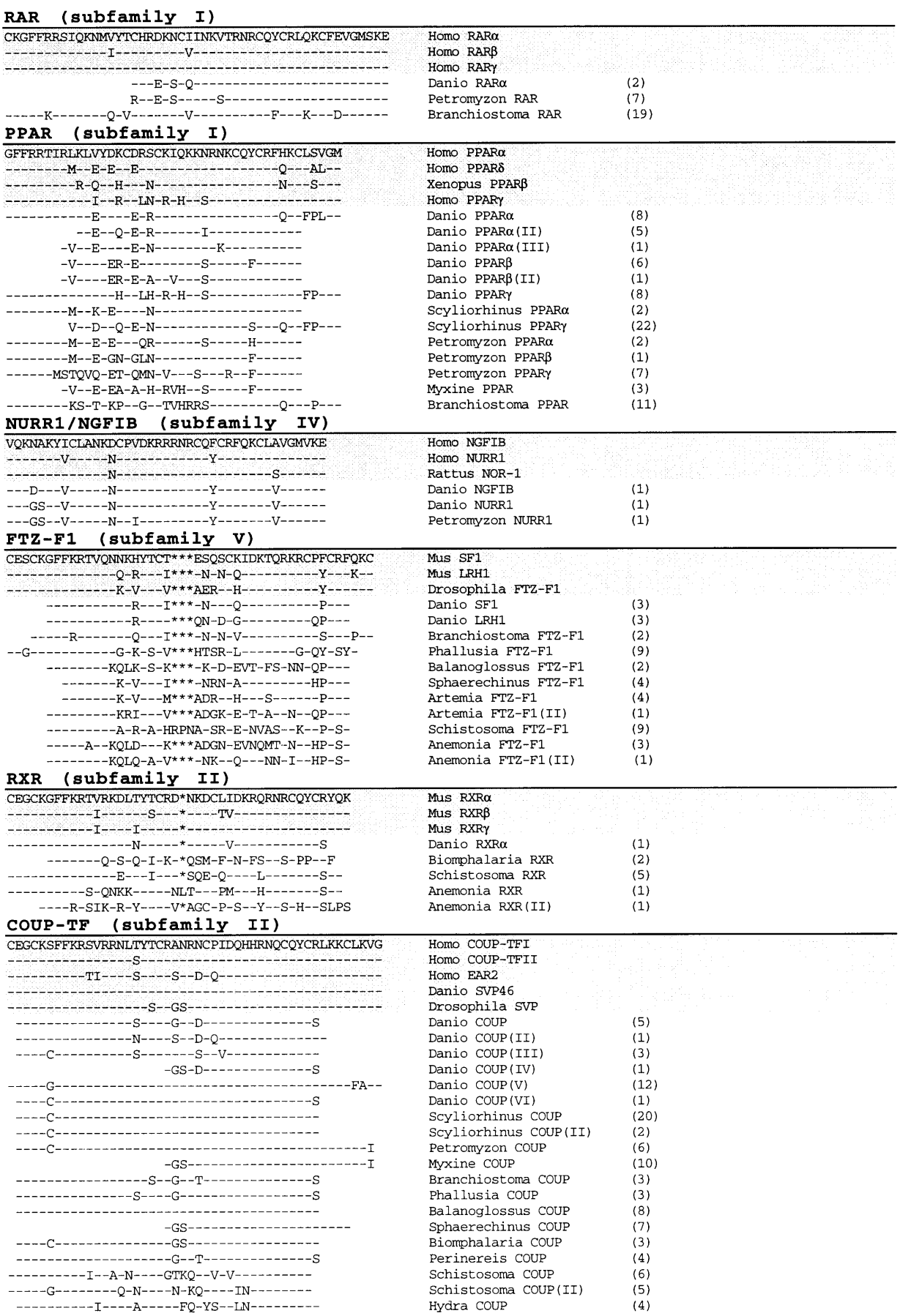

FIG. 1. Deduced amino acid sequence alignment of the NR PCR products corresponding to 6 of the 13 studied groups. The sequences are compared with their mammalian or arthropod homologues, which are shaded. Only amino acids different to those in the first sequence are indicated.

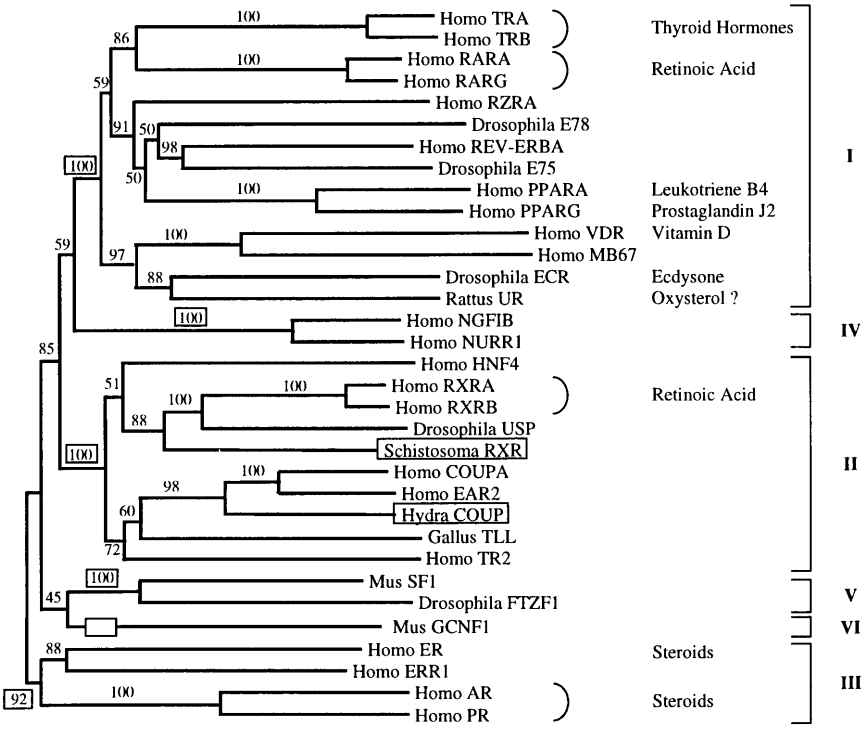

FIG. 2. Phylogenetical tree of 33 selected NRs. The six subfamilies are shown. Subfamilies are defined by clusters supported by high robustness "bootstrap" values above 90 starting from the base of the tree, which was rooted at mid-length of its longest path. Ancestral segments for subfamilies 1 to 5 have their bootstrap value boxed. Subfamily VI defined by only one member is shown by a small box along its branch. Ligands are indicated when applicable. The two boxed sequences correspond to cDNA clones of early metazoan NRs isolated in this study and corresponding to the PCR products of Fig. 1.

handled. To avoid crosscontamination, the DNA from chordates and vertebrates was extracted and submitted to PCR after completion of the PCR done on non- and early-metazoan samples.

Cloning and Sequencing. PCR fragments were cloned in the TA cloning vector (Invitrogen) and sequenced using an Applied Biosystems 373A automatic sequencer and methods and reagents of the supplier.

Sequence Analysis. Assignment of each sequence to a particular group was done after removal of the primer sequence and translation by alignment and construction of phylogenetic trees. The assignment of the sequence of the PCR products corresponding to genes, which were then isolated as cDNA (see Table 1), was proven to be correct, suggesting that no major assignment errors were made using short sequences. Nevertheless, in some cases such as *Danio* (a partially tetraploid organism) the precise assignment to a given paralogous gene could be problematic.

Sequences of all known NR from databases (see ref. 1) were aligned using the MUST package (8). Only the C and E domains were analyzed. Distance trees were calculated using the neighbor-joining method with 1,000 bootstrap replicates. To decrease calculation times three independent sets of 33 receptors that gave identical results were treated separately. The clustering into subfamilies and groups was confirmed by parsimony analysis using the PAUP program (9). The complete phylogenetic reconstruction will be published elsewhere (V.L., unpublished work).

RESULTS AND DISCUSSION

Ten sets of primers were designed to amplify the second exon encoding the highly conserved DNA-binding domain of NRs outside the animal kingdom (in plants, algae, and fungi) as well as in sponges, cnidarians, acoelomate, and coelomate triploblasts, and in key organisms of the transition from chordates to vertebrates. By two successive rounds of PCR cycles using all possible combinations of nested primers and a "touchdown" PCR method (see *Materials and Methods*), we amplified 84 similar genomic DNA fragments from a wide range of metazoans. We report here the first, to our knowledge, NR clones in early metazoans, such as cnidarians and acoelomates. Sponge DNA yielded no PCR fragments harboring a clear NR signature. The question of the presence of NR in this phylum, the phylogenetic position of which within metazoans is unclear, thus remains open (10, 11). Interestingly, we did not obtain any positive amplification outside metazoans. Although we cannot formally exclude the possibility that our primers may have artifactually missed NR sequences in these organisms, our observation confirms the apparent absence of NR observed in yeast (12). Thus, NRs appear to be confined to metazoans.

The NR gene fragments were cloned and sequenced (Fig. 1). Because the PCR method is particularly prone to contamination artifacts, we performed DNA extraction, PCR experiments, and cloning/sequencing in three different laboratories. Furthermore, the authenticity of the sequenced PCR products was ascertained by the following criteria (Table 1): (*i*) the A/T content of the PCR products was in accordance with the A/T

For each PCR product the number of sequenced clones is indicated. Gaps in the sequence alignment are indicated by stars. Interestingly, in *Anemonia* and *Schistosoma* we found several PCR products assigned to a given NR group (e.g., *Schistosoma* COUP and COUP II). In these cases the Roman number in parentheses is an arbitrary number needed to discriminate different gene versions. This suggests that, at least in these organisms, there is more diversity in NR than previously expected. The precise position of these extra copies was not determined, and it is not known whether these are secondary duplications specific to a given species or rather new receptor genes. In all cases these duplications are independent of the vertebrate gene duplication discussed in Fig. 3. GenBank codes of the sequences are U93406 to U93489.

FIG. 3. Simplified phylogeny of metazoans (*A*) and chordates (*B*) showing the organisms where NR genes were found. In *A*, the subfamilies are indicated for each sequence. For *B* the number of different sequences found for a given group (e.g., PPAR) is indicated in parentheses. Zoological groups are boxed, previously described receptors from the literature are in bold, and receptors identified in this study are italicized.

content of the genomes in the case of *Hydra* and *Schistosoma* (13); (*ii*) Southern blots using the PCR products as probes were done when sufficient amounts of DNA were available and revealed specific bands under high-stringency conditions; (*iii*) cDNA and genomic libraries were screened using the PCR products and revealed clones containing a region 100% identical with the probe; (*iv*) several independent PCR amplifications in the same organisms produced identical sequences; and (*v*) the position of the fragments in phylogenetic trees was consistent with their origin (14). Taken together these criteria provide strong evidence that the amplified DNAs correspond to bona fide NR gene sequences in these various organisms.

The NR sequences were then compared with other NRs (Fig. 1). Interestingly, in cnidarians and *Schistosoma* we identified only members of COUP-TF, RXR, and FTZ-F1 groups of receptors. Despite an extensive search, no members of the other groups (such as RAR, TR, etc.) were found in these animals. Importantly, these primers have been successfully tested in a wide range of organisms that appeared more recently in evolution in which they give positive results. In addition, the NRs we found to be missing in cnidarians and *Schistosoma* are not evolving more rapidly in vertebrates than the sequences we effectively found. To position COUP-TF, RXR, and FTZ-F1 sequences, we constructed a phylogenetic tree connecting a representative selection of full-length NR sequences (Fig. 2). By observing the confidence "bootstrap" values of the most internal branches of the tree, the superfamily can be separated into six subfamilies and 26 groups of receptors. The relationships among the subfamilies remain unresolved. Nevertheless, the fact that the groups in which we found homologues in early metazoans only correspond to two subfamilies (II and V, Figs. 1 and 2) suggests that the origin of the superfamily can be positioned between these two subfamilies. Furthermore, the placement of the full-length sequence of the *Hydra* COUP cDNA that we recently isolated shows that *Hydra* COUP diverged before the split that led to COUP-TF and EAR2 and after the appearance of TLL and TR2 (Fig. 2). This indicates that a first wave of gene duplication took place before cnidarians diverged from bilaterians.

A second wave of gene duplications can be observed on the tree that led to the diversification of receptors inside each group, i.e., the appearance of the two to four paralogous copies of each receptor type (RARA, RARB, and RARG, etc.). Within the Deuterostomians, this wave of gene duplication is specific to vertebrates because in all cases we observed only one homologue in tunicates or amphioxus (Figs. 1 and 3*B*). This homologue (e.g., amphioxus COUP, RAR, or PPAR) is located as a nonduplicated version of the paralogous vertebrate receptors (data not shown). In contrast, in early vertebrates such as lamprey, shark, or zebrafish we found several paralogous genes. This means that orthologues of each gene (e.g., RARA, RARB, and RARG) in higher vertebrates are found in early vertebrates, hence the paralogous groups evolved before the latest common ancestor of living vertebrates (Fig. 1 and data not shown). In this sense, the four steroid receptors (PR, GR, MR, and AR) should be viewed as vertebrate-specific paralogous copies of a unique ancestral steroid receptor. Because NR genes are dispersed throughout the genome, our observation is clearly in favor of the Ohno model of vertebrate specific duplication of the entire genome (15, 16). A large number of new genes that were able to play a crucial role in the appearance of the vertebrate-specific developmental innovations were created by this event (15).

In contrast to orphan receptors (such as REV-ERBA, NGFIB, or SF1), vertebrate receptors with ligands have, in most cases, no arthropod homologues (Fig. 2), suggesting that orphan receptors are ancient (i.e., primitive), whereas receptors to known ligands are modern (i.e., derived). Indeed, receptors to known ligands appear to be present only in coelomates (Figs. 2 and 3*A*). In this context, the case of RXR is particularly interesting, because it has a homologue, called USP, in arthropods. Interestingly, the function of this homologue appears to be different in terms of ligand binding. In fact, RXR binds a ligand (9-*cis*-retinoic acid) in vertebrates but does not do so outside vertebrates, as exemplified by the *Drosophila USP* gene product (17). It is thus conceivable that there was a gain of 9-*cis*-retinoic acid binding to RXR specifically in vertebrates. Thus, the RXR/USP homologues that we found in early metazoans (*Anemonia* and *Schistosoma*) should be orphan receptors. In accordance with the above hypothesis of acquisition of ligand binding function, RAR and RXR, which both bind retinoids, are located on different subfamilies in our phylogenetic tree. Strikingly, we noticed that, in general, there is no relationship between the position of a liganded NR in the tree and the chemical nature of its ligand. Thus, because the liganded receptors appear to be predominantly members of recent subfamilies of receptors, we propose that they have gained the ability to bind their ligands independently and that the ancestral NR was an orphan receptor. This model has important functional implications because it suggests that the conformational change that governs the activity of the receptor can be achieved in the absence of a ligand binding. Recently, it has been proposed that NRs do not exist in static off or on conformations but that ligand alters an equilibrium between the inactive and active states (18). Acquisition of ligand binding during the course of evolution could be an easy way to lock the conformation of the receptor into an active state. Indeed, the recent demonstration that the activation domain AF-2 of NRs such as RXR or PR may be active (i.e., providing binding surfaces for coactivators) in the absence of ligand (e.g., by phosphorylation) is in perfect accordance with our model (19–21).

Acquisition of ligand binding function within a transcription factor superfamily is not unique to NRs. For example, a member of the basic helix–loop–helix family has gained the ability to bind dioxin (22). Similarly, an independent gain of ligand binding ability also was suggested for the G-protein-coupled bioamine receptors in which divergent groups of receptors bind related ligands (exactly like RAR and RXR in NRs) whereas highly related groups bind structurally different molecules (as do TR and RAR in NRs) (23). Future structural and functional studies addressing the three-dimensional organization, pattern of expression, ligand-dependent structural and functional alterations, and dimerization abilities of early-metazoan NR homologues will cast a new light on how the multiple functions present in these molecules evolved and will allow the definition of the role of these genes in the evolutionary process itself.

We thank Laurent Meijer for access to laboratory facilities; Blandine Baratte, Murielle Fikry-Vennin, and Agnès Bègue for technical help; Brigitte Galliot, Jean-Guy Catford, François Fontaine, Robert Garrone, and Guillaume Lecointre for gifts of samples; and Hinrich Gronemeyer, François Catzeflis, Philippe Berta, Ricardo Mendonça, Jean Coll, and Jean Marc Vanacker for critical reading of the manuscript. We thank the anonymous reviewers for their comments. H.E. received support from the European Community (Human Capital and Mobility Program) and from Fondation Les Treilles. M.C.L. was funded by Association pour la Recherche Contre le Cancer. This work was supported by grants from Groupement de Recherche et D'etude Desgenomes, Centre National de la Recherche Scientifique, Institut National de la Santé et de la Recherche Médicale, and Institut Pasteur de Lille.

1. Gronemeyer, H. & Laudet, V. (1995) *Protein Profile* **2,** 1173–1308.
2. Mangelsdorf, D. J. & Evans, R. M. (1995) *Cell* **83,** 841–850.
3. Laudet, V., Hänni, C., Coll, J., Catzeflis, F. & Stéhelin, D. (1992) *EMBO J.* **11,** 1003–1013.
4. Amero, S. A., Kretsinger, R. H., Moncrief, N. D., Yamamoto, K. R. & Pearson, W. R. (1992) *Mol. Endocrinol.* **6,** 3–7.
5. Detera-Wadleigh, S. D. & Fanning, T. G. (1995) *Mol. Phylogenet. Evol.* **3,** 192–205.
6. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, New York), 2nd Ed.
7. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991) *Nucleic Acids Res.* **19,** 4008.
8. Philippe, H. (1993) *Nucleic Acids Res.* **21,** 5264–5272.
9. Swofford, D. L. & Olsen, G. J. (1990) in *Molecular Systematics*, eds. Hillis, D. M. & Moritz, D. (Sinauer, Sunderland, MA), pp. 411–501.

10. Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K. G. & Adoutte, A. (1991) *EMBO J.* **10,** 499–503.
11. Cavalier-Smith, T., Allsopp, M. T. E. P., Chao, E. E., Boury-Esnault, N. & Vacelet, J. (1996) *Can. J. Zool.* **74,** 2031–2045.
12. Dujon, B. (1996) *Trends Genet.* **12,** 263–267.
13. Musto, H., Alvarez, F., Tort, J. & Rodriguez-Maseda, H. (1994) *Int. J. Parasitol.* **24,** 277.
14. Philippe, H., Chenuil, A. & Adoutte, A. (1994) *Development (Cambridge, U.K.),* Suppl., 15–25.
15. Holland, P. W. H., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. (1994) *Development (Cambridge, U.K.),* Suppl., 125–133.
16. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Heidelberg).
17. Mangelsdorf, D. J., Borgmeyer, U., Heyman, R. A., Zhou, J. Y., Ong, E. S., Oro, A. E., Kakizuka, A. & Evans, R. M. (1992) *Genes Dev.* **6,** 329–341.
18. Schulman, I. G., Juguilon, H. & Evans, R. M. (1996) *Mol. Cell. Biol.* **16,** 3807–3813.
19. Matkovits, T. & Christakos, S. (1995) *Mol. Endocrinol.* **9,** 232–242.
20. Perlmann, T. & Vennström, B. (1995) *Nature (London)* **377,** 387–388.
21. Power, R. F., Mani, S. K., Codina, J., Connelly, O. M. & O'Malley, B. W. (1991) *Science* **254,** 1636–1639.
22. Burbach, K. M., Poland, A. & Bradfield, C. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8185–8190.
23. Vernier, P., Cardinaud, B., Valdenaire, O., Philippe, H. & Vincent, J. D. (1995) *Trends Pharmacol. Sci.* **16,** 375–376.